

**APPARATUS FOR ANALYSING CARDIAC EVENTS*****Technical field***

The present invention relates to an apparatus for analysing cardiac events  
5 detected in electrograms, EGMS, and to a heart stimulator provided with such an apparatus.

In the following the expression cardiac event denotes the depolarization phase in the cardiac cycle which for atrial signals is commonly known as P wave and for ventricular signals as R wave or QRS complex.

10

***Background***

In the field of devices for cardiac rhythm management (CRM), accurate rhythm classification is an increasingly important aspect. Pacemakers are primarily used to assist in bradycardia or when the electrical propagation path is blocked,  
15 whereas the primary use of implantable cardioverter defibrillators (ICD) is to terminate ventricular arrhythmia, a life-threatening condition if not acutely treated. In both types of devices, accurate event classification of the electrogram signal is needed for identifying, e.g., atrial and ventricular fibrillation in order to give appropriate therapy for the detected arrhythmia. For pacemakers, this may imply  
20 changing the pacing mode in order to stabilize the ventricular rhythm during an episode of atrial fibrillation. An ICD responds to ventricular fibrillation by giving a defibrillatory shock intended to terminate the fibrillation.

Ever increasing demands are put on both kinds of devices to better handle their primary task as well as to manage other tasks than those originally intended for. One such task may, for an implantable medical device, be to identify atrial flutter in order to terminate it by atrial pacing or to defibrillate atrial fibrillation.  
25 Although it is not a life-threatening arrhythmia, atrial fibrillation is an inconvenience to the patient and increases the risk for other diseases such as stroke. Atrial pacing may also be one way of terminating supraventricular tachycardias. An ICD specific task is to identify atrial fibrillation in order to not mistake it for ventricular  
30 fibrillation and the risk of giving an unnecessary, and possibly harmful, defibrillation shock. Another, more general, utilization is to efficiently store rhythm data for later analysis and evaluation, already done in modern ICD's. By collecting

data, better knowledge of the evolution of cardiac diseases and the functionality of the device can be obtained.

Clustering represents an important task within the classification problem where each individual event is assigned to a cluster of events with similar features.

5 Labelling of the clusters, i.e., associating the cluster with a specific cardiac rhythm, completes the classification such that the device can provide proper therapy when needed. However, certain constraints distinguish clustering in CRM devices from clustering in general. In order to give immediate therapy, it requires clustering to be done in real-time, thus excluding many iterative clustering algorithms such as k-

10 means clustering and competitive learning. Various methods have recently been presented concerning clustering of signals from the surface electrocardiogram (ECG), based on, e.g., self-organizing maps or fuzzy hybrid neural networks, see M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrabdt, and L. Sörnmo, "Clustering ECG complexes using Hermite functions and self-organizing maps", IEEE Trans. Biomed. Eng., vol. 47, pp. 838 –848, July 2000, and S. Osowski and T. Linh, "ECG beat recognition using fuzzy hybrid neural network", IEEE Trans. Biomed. Eng., vol. 48, pp. 1265 –1271, November 2001. However, most clustering algorithms used for ECG analysis are computationally rather complex and therefore unsuitable for implantable CRM devices. Furthermore, not much a priori morphologic information is associated with the various rhythms in the electrogram (EGM); this is in contrast to the more well-defined ECG.

15

20

Previously presented work in the area of intracardiac event classification mainly focus on discrimination of a specific condition in order to discern, e.g., atrial fibrillation from other atrial tachyarrhythmias, see A. Schoenwald, A. Sahakian, and S. Swiryn, "Discrimination of atrial fibrillation from regular atrial rhythms by spatial precision of local activation direction", IEEE Trans. Biomed. Eng., vol. 44, pp. 958 – 963, October 1997. Other applications involve discrimination of ventricular from supraventricular tachycardia, see L. Koyrakh, J. Gillberg, and N. Wood, "Wavelet based algorithms for EGM morphology discrimination for implantable ICDs", in Proc. Of Comp. In Card. (Piscataway, NJ, USA), pp. 343 – 346, IEEE, IEEE Press, 1999, and G. Grönfeld, B. Schulte, S. Hohnloser, H. - J. Trappe, T. Korte, C. Stellbrink, W. Jung, M. Meesmann, D. Böcker, D. Grosse – Meininghaus, J. Vogt, and J. Neuzner, "Morphology discrimination: A beat-to-beat algorithm for the discrimination of ventricular from supraventricular tachycardia by implantable

25

30

cardioverter defibrillators", J. Pacing Clin. Electrophysiol., vol. 24, pp. 1519 – 1524, October 2001. More general classification algorithms, which in turn involve training on individual patients, have been based on analogue neural networks or wavelet analysis for morphologic discrimination of arrhythmias.

5 WO 97 39 681 describes a defibrillator control system comprising a pattern recognition system. The intracardiac electrogram signal is digitised and delivered for feature selection into a selector. The feature selector outputs selected features to a trained classifier to provide information as to what group the produced signal should be clustered, e.g. ventricular tachycardia. The classifier  
10 outputs the classified information for use for a therapeutic decision.

US 5 271 411 discloses an ECG signal analysis and cardiac arrhythmia detection by extraction of features from a scalar signal. A QRS pattern vector is then transformed into features describing the QRS morphology, viz. a QRS feature vector. A normal QRS complex is identified based on the population  
15 of QRS complexes located within clusters of QRS features within a feature space having a number of dimensions equal to the number of extracted features. The extracted morphology information is then used for judging whether a heart beat is normal or abnormal.

US 5 638 823 describes non-invasively detecting of coronary artery  
20 diseases. A wavelet transform is performed on an acoustic signal representing one or more sound event caused by turbulence of blood flowing in an artery to provide parameters for a feature vector. This feature vector is used as one input to neural networks, the outputs of which represent a diagnosis of coronary stenosis in a patient.

25 In Michael A. Unser et al, "Wavelet Applications in Signal and Image Processing IV", Proceedings SPIE – The International Society for Optical Engineering, 6 – 9 August 1996, vol. 2825, part two of two parts, pp. 812 – 821, a wavelet packet based compression scheme for single lead ECGs is disclosed, including QRS clustering and grouping of heart beats of similar structures. For  
30 each heart beat detected, its QRS complex is compared to templates of previously established groups. Point-by-point differences are used as similarity measures. The current beat is assigned to the group whose template is most similar, provided predetermined conditions are satisfied. Otherwise a new group is created with the current QRS complex used as the initial group template.

Future pacemakers and ICD's will include more advanced means for arrhythmia detection and therapy, and the purpose of the present invention is to propose a technique for separation of cardiac rhythms in a reliable way on the  
5 basis of electrogram event clustering

***Disclosure of the invention***

The above purpose is obtained by an apparatus according to claim 1.

Certain arrhythmias are diagnosed immediately to give proper therapy,  
10 while, for others, it may be sufficient to record the rhythm for data collection purposes. Thus, the classification problem, viz. to label the rhythms based on clusters using clinical terms, may not always be necessary to implement.

According to advantageous embodiments of the apparatus according to the invention both morphologic and temporal data are considered for clustering.  
15 Morphologic features are efficiently extracted by use of the dyadic wavelet transform after which the events are grouped by a leader-follower clustering embodiment. The event detection problem, based on the same transform, is previously treated in Swedish patent application no. 0103562-5.

According to another advantageous embodiment of the apparatus  
20 according to the invention an integrating means is provided to integrate said distance over a predetermined period of time. By integrating the distance over a period of time it is possible to distinguish irregular rhythms, like atrial fibrillation, from regular rhythms. The integral total distance in case of atrial fibrillation will be high whereas regular rhythms will result in a lower total distance.

25 The invention also relates to a heart stimulator provided with the above mentioned apparatus for controlling the therapeutic stimulation depending on arrhythmia detection.

***Brief description of the drawings***

30 To explain the invention in greater detail embodiments of the invention, chosen as examples, will now be described with reference to the enclosed drawings, on which figure 1 shows impulse responses of a filter bank used for cardiac event detection, figure 2 is a flow chart of clustering algorithm perform with the apparatus according to the invention, figure 3 illustrates the computational

complexity for different clustering algorithms, figure 4 presents in (a) clustering performance in terms of probability of a correct clustering of an event,  $P_{CC}$  and probability of a dominant event in a cluster  $P_{DE}$  and in (b) the spread in the number of clusters, figure 5 illustrates in (a) clustering performance in terms of  $P_{DE}$  and

- 5      P<sub>CC</sub> as a function of a distance threshold  $\eta$  and in (b) the numbers of clusters as a function of  $\eta$ , figure 6 exemplifies the clustering performance for one set of parameters, figure 7 exemplifies clustering result on a concatenated EGM for three cases, figure 8 is a flow chart illustrating the function in broad outline of an embodiment of the apparatus according to the invention, and figure 9 is a block
- 10     diagram of an exemplifying embodiment of a heart stimulator provided with an apparatus according to the invention.

### ***Description of preferred embodiments***

#### ***P wave detection and feature extraction***

- 15     A signal model assuming that the event waveform is composed of a linear combination of representative signals is considered. The feature extraction problem is then to estimate the individual components of the representative signals since each morphology will have its own linear combination. By using the dyadic wavelet transform, different widths of the two fundamental monophasic and
- 20     biphasic waveforms are included in the model at a low cost.

#### ***Feature extraction***

It is assumed that the QRS waveform is composed of a linear combination of representative signals,

$$25 \quad H = [h_1, \dots, h_p] \quad (2)$$

where each function,  $h_j$ ,  $j = 1, \dots, P$ , is of size  $M \times 1$ . The only restriction on  $H$  is that it must have full rank. Different morphologies,  $s(n)$ , are modelled by the  $P \times 1$  coefficient vector  $\theta(n)$ , with the linear model

$$s(n) = H\theta(n) \quad (3)$$

- 30     where  $n$  is a temporal variable describing when the event occurs. The observed signal,  $x(n)$ , is assumed to be modelled by,

$$x(n) = H\theta(n) + w(n) \quad (4)$$

where the  $M \times 1$  noise vector  $w(n)$  is assumed to be zero-mean, white, and Gaussian with variance  $\sigma_w^2$ . Consequently,  $x(n)$  is defined as

$$x(n) = \begin{bmatrix} x(n) \\ \vdots \\ x(n+M-1) \end{bmatrix} \quad (5)$$

implying an event duration of  $M$  samples, beginning at  $n$ .

5 function  $x(n)$  for a specific realization of  $\theta(n)$ ,  $p(x(n); \theta(n))$ , is thus given by

$$p(x(n); \theta(n)) = \frac{1}{(2\pi\sigma_w^2)^{\frac{M}{2}}} \exp \left[ -\frac{1}{2\sigma_w^2} (x(n) - H\theta(n))^T (x(n) - H\theta(n)) \right] \quad (6)$$

In this model, a complete description of a QRS complex is provided by the deterministic unknown parameter vector  $\theta(n)$ . The absence of a QRS complex

10 corresponds to the case when  $\theta(n)$  is equal to 0, where 0 is the  $P \times 1$  zero vector. In general, no *a priori* knowledge is available on  $\theta(n)$ , and therefore an estimate is required before detection can take place. Furthermore, only one event is assumed to take place within the observation interval  $0 \leq n \leq N$ .

## 15 *Filterbank representation*

The descriptive functions in  $H$  have been selected such that the following three aspects have been taken into particular account:

- 1. the main morphologies of the QRS complex are mono- and/or biphasic
- 2. the broad range of QRS complex durations, and
- 20 3. a low complexity implementation.

The wavelet transform is particularly suitable since it is a local transform, i.e., it provides information about the local behaviour of a signal. One wavelet decomposition method which may be efficiently implemented is the dyadic wavelet transform. By careful selection of the filters, a suitable filter bank including mono-  
25 and biphasic impulse responses can be obtained. A symmetric lowpass filter,  $f(n)$ , is used repeatedly in order to achieve proper frequency bands. This filter is combined with one of two filters,  $g_b(n)$  or  $g_m(n)$ , which together define the

waveform morphology (the subindices b and m denote bi- and monophasic, respectively). For the biphasic case, the recursion is expressed as,

$$\begin{aligned}
 h_{1,b}(n) &= g_b(n) \\
 h_{2,b}(n) &= f(n) * g_b(2n) \\
 5 \quad h_{3,b}(n) &= f(n) * f(2n) * g_b(4n) \\
 &\vdots \\
 h_{q_{\max},b}(n) &= f(n) * \dots * f(2^{q_{\max}-3}n) * g_b(2^{q_{\max}-1}n)
 \end{aligned} \tag{7}$$

in which the subindex  $q_{\max}$  represents the maximum (coarsest) scale.

10 It is now possible to present an expression for H which is composed of one biphasic and one monophasic part,

$$H = \begin{bmatrix} \tilde{H}_b & \tilde{H}_m \end{bmatrix} \tag{8}$$

15 where the biphasic  $H_b$  is defined as

$$H_b = \begin{bmatrix} h_{q_{\min},b}(0) & \dots & h_{q_{\max},b}(0) \\ \vdots & \ddots & \vdots \\ h_{q_{\min},b}(M-1) & \dots & h_{q_{\max},b}(M-1) \end{bmatrix} \tag{9}$$

where the subindex  $q_{\min}$  represents the minimum scale and  $q_{\min} < q_{\max}$ . The  
20 monophasic matrix  $H_m$  is computed in a corresponding way. The reversed order of  
the columns in  $\hat{H}$ , denoted with  $\tilde{H}$  in (8), is introduced in order to be consistent with  
the model assumed in (4).

In order to mimic the desired mono- and biphasic waveforms with a  
low complexity filter bank structure, short filters with small integer coefficients were  
25 used. In (7), the impulse response  $f(n)$  was chosen as a third order binomial  
function,

$$F(z) = (1+z^{-1})^L \stackrel{L=2}{=} 1 + 3z^{-1} + 3z^{-2} + z^{-3} \tag{10}$$

where  $F(z)$  is the Z-transform of  $f(n)$ . For the biphasic filter bank, the  
30 filter  $g_b(n)$  was selected as the first order difference,

$$g_b(n) = [-1 \ 1] \quad (11)$$

The filter  $g_m(n)$  was chosen such that a compromise between the requirement of a DC gain equal to zero and an approximately monophasic impulse response was achieved. A reduction of complexity results when  $g_b(n)$  is reused,

$$5 \quad g_m(n) = g_b(n) * g_b(n) = [1 -2 1] \quad (12)$$

For this particular choice of  $g_m(n)$  and by using Mallat's algorithm, it is possible to calculate both the biphasic and the monophasic filter output from each scale by using  $f(n)$  once and  $g_b(n)$  twice. The filter bank impulse responses are shown in Fig. 1. The filter bank includes two orthogonal signal sets where the 10 width of the signal varies within each set. In (a),  $h_{j,b}(n)$  is shown for  $j = 2, \dots, 4$  from top to bottom, and in (b), the corresponding  $h_{j,m}(n)$  are shown.

### **ML parameter estimation**

The unknown coefficient vector  $\theta(n)$  can be estimated by using the maximum likelihood criterion according to,

$$15 \quad \hat{\theta}(n) = \arg \max_{\theta(n)} p(x(n); \theta(n)) \quad (13)$$

However,  $\hat{\theta}(n)$  is only of interest for those  $n$  for which the probability of an event, or, equivalently, for which the likelihood ratio test function  $T(x(n))$  is maximized,

$$\hat{n} = \arg \max_n T(x(n)) \quad (14)$$

20 For this case,  $T(x(n))$  can be shown to be [14],

$$T(x(n)) = x(n)^T H \left( H^T H \right)^{-1} H^T x(n) \quad (15)$$

for the case when  $\sigma_w^2$  is assumed to be constant.

The optimal estimate  $\hat{\theta}$  for the detected event at  $\hat{n}$  is thus expressed as

$$\hat{\theta} = \arg \max_{\theta(\hat{n})} p(x(\hat{n}); \theta(\hat{n})) \quad (16)$$

25 The MLE of  $\theta(\hat{n})$  is found by maximizing  $p(x(\hat{n}); \theta(\hat{n}))$ , or equivalently minimizing the MSE,

$$(x(\hat{n}) - H\theta(\hat{n}))^T (x(\hat{n}) - H\theta(\hat{n})) = x(\hat{n})^T x(\hat{n}) + \theta(\hat{n})^T H^T H \theta(\hat{n}) - 2\theta(\hat{n})^T H^T x(\hat{n}) \quad (17)$$

Derivation with respect to  $x(\hat{n})$  yields the optimum  $\hat{\theta}$ ,

$$\hat{\theta}(\hat{n}) = \left( H^T H \right)^{-1} H^T x(\hat{n}) = \left( H^T H \right)^{-1} H^T x(\hat{n}) \quad (18)$$

By using the above formulation, it is also possible to derive a generalized likelihood ratiodetector based on (15).

#### **Rate**

5 The central parameter for classification of cardiac arrhythmias is the heart rate. Most arrhythmias are defined in terms of heart rate, although sometimes with rather fuzzy limits. Consequently, rate should be considered in order to improve performance. The RR interval,  $\Delta t$ , is defined as the duration between two consecutive events,

10 
$$\hat{\Delta t}_k = (\hat{n}_k - \hat{n}_{k-1})T_s \quad (19)$$

where  $\hat{n}_k$  and  $\hat{n}_{k-1}$  denote the occurrence times of the events, and  $T_s$  denotes the sampling period.

#### **Leader-follower clustering**

15 The choice of leader-follower clustering is based on a number of features which makes it suitable for the present invention, viz.

- o on-line processing (non-iterative), and
- o self-learning, i.e., no a priori knowledge of the of clusters is needed.

The starting point is the assumption that an event is present for which it  
20 should be decided whether it belongs to an already existing cluster or if a new cluster should be initiated. Since, no knowledge is a priori available on which rhythms or morphologies to be expected, the chosen algorithm must be self-learning. The leader-follower clustering algorithm is constituted by four quantities:

1. The event parameter vector,  $\Phi(n_k)$ , containing the features of the k:th event that occurs at time  $n_k$ .
2. The cluster center,  $\mu_i$ , and the covariance matrix  $\Sigma_i$  that together define the i:th cluster. Since both  $\mu_i$  and  $\Sigma_i$  are unknown, they are replaced by their estimates  $\hat{\mu}_i$  and  $\hat{\Sigma}_i$ , respectively.
3. The metric,  $d_i^2$ , determine the distance between  $\Phi(n_k)$  and  $\mu_i$  according to some suitable function.
4. A rule for adaptation of the cluster parameters for the winning cluster.

## Initialization of new clusters

During run-time, a finite number of clusters exists which represent the rhythms having appeared until present time. Thus, it is occasionally necessary to initialize a new cluster when the existing ones do not sufficiently well fit the present event. When the distance function  $d_i^2$  exceeds a certain threshold,  $\eta$ , it is more likely that the event belongs to a new cluster than to any of the existing clusters. The selection of  $\eta$  is a tradeoff between cluster size and cluster resolution in the sense that choosing a small  $\eta$  will result in many clusters with few clustering errors. On the other hand, a large  $\eta$  results in few clusters but in more errors.

10 The minimum distance between  $\Phi(n)$  and  $\mu_i$  with respect to both  $i$  and  $n$  is

$$d_{\min}^2 = \min_{i,l} d_i^2(l) \quad i=1,\dots,I \quad l=\hat{n}_k - \frac{K}{2}, \dots, \hat{n}_k + \frac{K}{2} \quad (20)$$

15 over the search interval K. The corresponding minimum distance  
indices are found as

$$[i_{\min}, n_{\min}] = \arg \min_{i,j} d_i^2(l) \quad (21)$$

The decision rule based on the comparison of  $d_{\min}^2$  and  $\eta$  is expressed as

20

$d_{\min}^2$  } > $\eta$  : Initialize new cluster  
           < $\eta$ : Assign to winning cluster  
           < $\rho\eta$ : Update winning cluster (22)

where  $0 < \rho < 1$ . When the upper relation in (22) holds, a new cluster is initialized by first increasing the number of clusters by one,  $I = I + 1$ , given that  $I < I_{\max}$  where  $I_{\max}$  is the maximum number of clusters, and then initializing the new cluster as.

$$\hat{\mu}_1 = \Phi(\hat{n}_k) \quad (23)$$

On the other hand, if  $I = I_{\max}$  the algorithm needs to discard one of the existing clusters. This can be done by elimination of, e.g., the oldest or the

"smallest" cluster. By the term "smallest" cluster is meant that cluster which most rarely is fitted with a detected cardiac event.

For the middle and lower conditions in (22), the existing minimum distance cluster  $i_{\min}$  is selected as the winning cluster. However, only the lower condition

5 results in a cluster parameter update. The reason for including such a distinction is that only closely similar events should be used for cluster updates in order to reduce contamination.

#### ***Mahalanobis distance function***

10 The distance between the feature vector  $\theta(n)$  and each cluster  $\hat{\mu}_i$  is defined as the Mahalanobis distance, which is a normalized Euclidean distance in the sense that it projects the parameter vector elements onto univariate dimensions by including the inverse covariance matrix  $\Sigma_i^{-1}$ . Thus, a feature with a larger variance in  $\Phi(n)$  will be assigned a larger share of the hyperspace before normalization compared to that with a lower variance. A consequence of normalization is that the Mahalanobis distance works well on correlated data since  $\Sigma_i^{-1}$  then acts as a decorrelator.

15

When searching for the minimum distance, a grid search over  $n$  is performed. This grid search is necessary since it not only minimizes the distance but also results in a more accurate fiducial point estimate than what would be the case when only considering  $T(x(\hat{n}_k))$ . The minimum distance is thus found by a grid search with respect to all existing clusters,  $i = 1, \dots, I$ , and all feature vectors within the duration of an event  $I$  as defined in (20),

$$d_i^2(I) = (\Phi(I) - \hat{\mu}_i)^T \hat{\Sigma}_i^{-1} (\Phi(I) - \hat{\mu}_i) \quad (24)$$

25

#### ***Reference feature adaptation***

In order to track changes in the features of the different rhythms, adaptation of both  $\hat{\mu}_{i_{\min}}(k)$  and  $\hat{\Sigma}_{i_{\min}}^{-1}(k)$  are desirable; here the event index  $k$  has been included for clarity. For  $\hat{\mu}_{i_{\min}}(k)$ , an exponentially updated average is used:

$$\hat{\mu}_{i_{\min}}(k) = \hat{\mu}_{i_{\min}}(k-1) + \gamma \varepsilon(k) \quad (25)$$

where

$$\varepsilon(k) = \Phi(n_{min}) - \hat{\mu}_{i_{min}}(k-1) \quad (26)$$

5 The exponential update factor  $\gamma$  is confined to the interval  $0 < \gamma < 1$ .

The inverse covariance matrix,  $\hat{\Sigma}_{i_{min}}^{-1}(k)$ , is estimated by exponential averaging of the new cluster difference matrix  $\varepsilon(k) \varepsilon^T(k)$  using the update factor  $(1 - \alpha)$ ,

$$\hat{\Sigma}_{i_{min}}^{-1}(k) = \left( (\alpha) \hat{\Sigma}_{i_{min}}^{-1}(k-1) + (1-\alpha) \varepsilon(k) \varepsilon^T(k) \right)^{-1} \quad (27)$$

10

Using the matrix inversion lemma

$$A = B^{-1} + C D^{-1} C^T \quad (28)$$

$$A^{-1} = B - BC(D + C^T BC)^{-1} C^T B \quad (29)$$

15 and pairing the terms in (27) with the ones in (28),

$$A = \hat{\Sigma}_{i_{min}}^{-1}(k) \quad (30)$$

$$B^{-1} \alpha \hat{\Sigma}_{i_{min}}^{-1}(k-1) \quad (31)$$

$$C = \varepsilon(k) \quad (32)$$

20  $D^{-1} = (1 - \alpha) \quad (33)$

The inverse  $\hat{\Sigma}_{i_{min}}^{-1}(k-1)$  in (27) may be computed without any matrix inversions as,

$$\hat{\Sigma}_{i_{min}}^{-1}(k) = \alpha^{-1} \hat{\Sigma}_{i_{min}}^{-1}(k-1) - \frac{\alpha^{-1} \hat{\Sigma}_{i_{min}}^{-1}(k-1) \varepsilon(k) \varepsilon^T(k) \alpha^{-1} \hat{\Sigma}_{i_{min}}^{-1}(k-1)}{(1-\alpha)^{-1} + \varepsilon^T(k) \alpha^{-1} \hat{\Sigma}_{i_{min}}^{-1}(k-1) \varepsilon(k)} \quad (34)$$

25 By utilizing the matrix inversion lemma, the computational complexity of the operation is reduced from  $O(P^3)$  to  $O(P^2)$ .

### **Algorithm initialization**

In leader-follower clustering, clusters are initialized as they become needed. This feature is convenient since it does not automatically introduce any unused clusters. However, it also puts demands on the algorithm to be able to

5 create new clusters when necessary, and also to terminate clusters either not used for long or with only a few events. Initially, the total number of clusters,  $I$ , is equal to one. The algorithm is initialized by assigning the parameter vector  $\Phi(\hat{n}_1)$ , which maximizes the test statistic in (14) of the first event, to the first cluster, cf. (23),

10

$$\hat{\mu}_1 = \Phi(\hat{n}_1) \quad (35)$$

For the general case,  $\Phi(\hat{n}_k)$  is composed of a subset of the representative functions together with the preceding RR-interval,

15

$$\Phi(\hat{n}_k) = \begin{bmatrix} \hat{\theta}_s(\hat{n}_k) \\ \|\hat{\theta}_s(\hat{n}_k)\| \\ \Delta t_k \end{bmatrix} \quad (36)$$

Where  $\hat{\theta}_s(\hat{n}_k)$  is a subset of the most discriminating elements in  $\hat{\theta}_s(\hat{n}_k)$ .

Note that an event is defined by its depolarization wave. Consequently,  $\Delta t$  is not included in the arrival time estimation of  $\hat{n}_k$ , but instead computed

20 afterwards. The time continuous notation of  $\Delta t_k$  is preferred since it results in a suitable magnitude similar to the normalized morphological information in  $\theta_s$ .

The inverse correlation matrix estimate is initialized in the same way for all clusters; a simple solution is to set it equal to a scaling of the identity matrix  $I$ ,

$$\hat{\Sigma}_i^{-1} = \delta I \quad (37)$$

25 where  $\delta$  is a design parameter. The complete clustering algorithm for organized events is presented in the flow chart in Fig. 2.

### **Reduced-complexity clustering**

In order to develop a more efficient algorithm in terms of performance versus power consumption and to evaluate the power consumption itself, a simple measure of the computational complexity can be used, namely, the total number of

5 multiplications. This quantity represents a much more complex operation than do additions. In this algorithm, where most operations are of the nature "multiply-accumulate", the number of additions is of the same order as the number of multiplications and may thus be neglected without significant loss of accuracy.

In order to reduce the computational complexity, focus is put on reducing  
10 the number of multiplications. The dominant contributions of multiplication operations are found in (24) and (34) which require  $P(P + 1)$  and  $P/2(3P + 5)$  multiplications, respectively, considering certain symmetry properties of  $\hat{\Sigma}_i^{-1}$ .

Furthermore, one division is required in (34). However, according to (20), (24) is performed  $IK$  times per event while (34) is performed only once per event.

15 Based on the above performance figures, a few approximations can be identified:

- o to use only the peak(s) in  $T(x(\hat{n}))$  instead of a complete grid search,
- o to use a simplified  $\hat{\Sigma}_i^{-1}$ , and
- o to use a likelihood based search sequence over the clusters, i.e., to

20 start with the most likely cluster and to stop the search if a sufficiently small distance is found.

Simplifying the grid search from spanning both samples and clusters in (20) to span over only clusters,

$$d_{\min}^2 = \min_i d_i^2(\hat{n}) \quad (38)$$

25 the number of multiplications may be reduced by a factor  $K$  to  $IP(P + 1)$ . Due to sensitivity in  $T(x(\hat{n}))$ , the feature vectors resulting in the peaks for two different events may differ significantly, this simplification is likely to result in more clusters. A useful compromise may instead be to use the coefficients from, e.g., the 3 largest peaks in  $T(x(\hat{n}))$  resulting in  $3IP(P + 1)$  multiplications per event.

30 Since a cardiac event lasts for longer time than one sample those samples which give the filter coefficients which are most similar to the cluster

reference are determined. A grid search over 40 msec is preferably made. In this way coefficients of greatest importance locally are determined. For this decision  $T(x(n))$  is considered and samples generating a peak are chosen, i.e.  $T(x(n-1)) < T(x(n)) > T(x(n+1))$ , since they indicate the probability for the presence of an event.

5 Another simplification, based on the approximation of orthogonality between the different wavelet scales as well as the RR interval, is to simplify  $\hat{\Sigma}_i^{-1}$  by only including its diagonal elements in the adaptation,

$$\hat{\Sigma}_{i_{\min}}^{-1} \approx \alpha^{-1} \hat{\Sigma}_{i_{\min}}^{-1} + (1-\alpha) \text{diag}(\varepsilon(k) \varepsilon^T(k)) \quad (39)$$

where  $\text{diag}(A)$  returns the diagonal elements of the square matrix A. By using this  
10 approximation, the number of multiplications used for the estimation of  $\hat{\Sigma}_i^{-1}$  is reduced to  $3P$ . Additionally, the distance computation in (20) is simplified and may be reduced to  $2IKP$  multiplications per event.

15 A reasonable assumption is often that successive events originate from the same rhythms. Considering this knowledge, it would be sufficient to compute the distances for the previously selected cluster. In doing so, the number of multiplications in (38) are reduced even further.

Table 1 presents the different detector versions as defined by their distinguishing features and shows computational complexity for the different versions of the clustering algorithm.

20

Version	Features		
	$\hat{\Sigma}_{i_{\min}}^{-1}$	Search alg.	Complexity C
A	Full	Interval	$IKP(P + 1) + \frac{P}{2} (3P + 5)$
B	Diagonal	Interval	$2IKP + 3P$
C	Full	3 peak	$3IP(P + 1) + \frac{P}{2} (3P + 5)$
D	Diagonal	3 peak	$6IP + 3P$

The total computational complexity, C, as reflected by the number of multiplications for the different algorithm versions, is presented in Fig. 3 as a

function of  $I$ . In figure 3 the solid line shows an algorithm version A, dashed line a version B, dotted line a version C, and a version D by dash-dotted line), for a feature vector with (a),  $P = 4$ , and (b),  $P = 7$  elements.

## 5 *Results*

The results are obtained by studying the performance of the following quantities:

- o algorithm versions, as presented in Table 1, and
- o noise tolerance, for noise-free signals and for signals with background noise of  $50\mu V$  RMS.

10

The following parameter settings have been used (unless otherwise stated),

$$\alpha = 1.05 \quad \gamma = 0.025 \quad \delta = 50 \quad K = 40 \quad (49)$$

15

It is noted that  $\alpha^{-1}$  is chosen to offer faster adaption than  $\gamma$ . The reason for that is that the initial estimate  $\hat{\Sigma}_i^{-1}$  is likely to be less accurate than the initial estimate  $\hat{\mu}_i$ . Also,  $\delta$  is chosen to have the same order of magnitude as the steady state eigenvalues  $\hat{\Sigma}_i^{-1}$ .

20

It should be noted that the different algorithm versions are not fully comparable in terms of performance for a specific  $\eta$  due to the differences in distance computation. In versions B and D, where a diagonal  $\hat{\Sigma}_{i_{\min}}^{-1}$  is used, the lack of non-diagonal information results in a nonorthogonal distance which is larger than the orthogonal one. Since versions C and D make use of a limited search, the minimum distance found may differ from the global minimum distance for the event. Both these algorithmic differences imply an increase in clustering quality for a certain value of  $\eta$ , however, at the expense of more introduced clusters.

## 30 *Evaluation measures*

The two main quantities evaluated are clustering performance and computational complexity, see Fig. 3; these two quantities are in general

conflicting. In the evaluation, a cluster is assigned to that cardiac rhythm which contains the most events in the cluster. This rhythm is denoted as the dominant rhythm within the cluster. With respect to the dominant rhythm, the cluster is defined as a correct cluster, whereas it is erroneous to all other rhythms. If a 5 rhythm is found to be dominant in more than one cluster, such clusters are first merged in the performance evaluation. The number of events in the correct cluster which belongs to the  $i$ :th dominant rhythm is equal to  $N_D(i)$ , while the number of events belonging to any other false rhythm in the cluster is equal to  $N_F(i)$ . The number of events of the dominant rhythm which are not classified in a correct 10 cluster, i.e., missing, is equal to  $N_M(i)$ .

The performance of the algorithm is evaluated in terms of probability of a correct clustering of an event,  $P_{cc}(i)$ , and probability of a dominant event in a cluster,  $P_{DE}(i)$ . The first parameter is expressed as the share of correctly clustered events within a rhythm and is, using the above parameters, defined by

$$15 \quad P_{cc}(i) = \frac{N_D(i)}{N_D(i) + N_M(i)} \quad (41)$$

The second parameter may be expressed as the share of dominant rhythm events within a cluster, and is defined as

$$P_{DE}(i) = \frac{N_D(i)}{N_D(i) + N_F(i)} \quad (42)$$

For the case when a rhythm completely lacks a correct cluster,  $P_{DE}(i)$  20 is undefined; the rhythm is then excluded from subsequent statistical computations. Averaging  $P_{cc}(i)$  and  $P_{DE}(i)$  over all clusters results in the global performance measures  $P_{cc}$  and  $P_{DE}$ , respectively.

It should be pointed out that  $P_{cc}(i)$  and  $P_{DE}(i)$  reach their maximal value of 1 when  $\eta$  is sufficiently small such that the number of clusters equals the 25 total number of beats evaluated. This is, of course, a highly undesirable solution although performance, as expressed in (41) and (42), will be excellent. For this reason, the total number of clusters,  $I$ , is a crucial parameter to be considered. Here, the average  $\bar{I}$  is used together with the minimum and maximum number of clusters for a case,  $I_{min}$ , and  $I_{max}$ , respectively.

30 The power consumption of the algorithm is an important parameter which determines the pacemaker life span. In this study, power consumption is approximated by the computational complexity defined above as the total number

of multiplications. As shown, the computational complexity depends mainly on three parameters: the feature vector length, P, the search interval, K and the cluster size, I.

## 5 *Noise-free signal clustering performance*

Figure 4 presents clustering performance in terms of  $P_{DE}$  and  $P_{CC}$ , see Fig. 4 (a), and  $I_{min}$ ,  $\bar{I}$  and  $I_{max}$ , see Fig. 4 (b). Thus in figure 4(a) clustering performance is shown as depending on  $\bar{I}$  for  $P_{CC}$  (dark bars) and  $P_{DE}$  (bright bars) for noise-free signals. In figure 4(b), the spread between the different cases is 10 shown as, from left to right,  $I_{min}$ ,  $\bar{I}$  and  $I_{max}$ . The presented algorithm versions are found in Table 1 and three values of  $\bar{I}$  have been chosen for comparison; 3, 4 and 5. Versions A and B perform similarly for all three cases, and achieve  $P_{DE} = 1$  and  $P_{CC} = 1$  for  $\bar{I} = 4$  and  $\bar{I} = 5$ , respectively, by creating an acceptable number of clusters. However, it can be seen from Fig. 4 (b) that a large difference in the 15 number of initialized clusters between different cases is present. For version B, a slight increase in both  $\bar{I}$  and  $I_{max}$  is observed for  $\bar{I} = 4$ . However, this increase is more due to an unfortunate step-like behaviour in the results for the given  $\bar{I}$  than for any significant decrease in performance. The values of  $\eta$  used in Fig. 4 for the different versions are shown in Table 2.

20 Table 2: Values of  $\eta$  used in Fig. 4.

$\bar{I}$	Algorithm version			
	A	B	C	D
5	3.6	4.2	5.4	5.8
4	4.6	4.8	7.0	7.2
3	9.6	9.6	12.0	10.8

Versions C and D perform slightly worse than do versions A and B. Contrary to the latter ones, neither version achieves  $P_{DE} = 1$  or  $P_{CC} = 1$  for the presented  $\bar{I}$ , see figure 4(a).

25 Figure 5 (a) presents the clustering performance for version A in detail and its dependence on  $\eta$ . The most noteworthy result is that both  $P_{DE} = 1$  and  $P_{CC} = 1$  for  $\eta < 7$ . It is also clear that clustering performance deteriorates rapidly for  $\eta > 10$ . The increase in  $P_{CC}$  for very high values of  $\eta$  is due to that not all rhythms are allocated to a dominant cluster and are therefore disregarded in the performance 30 computation.

In Fig. 5 (b),  $I_{\min}$ ,  $\bar{T}$  and  $I_{\max}$  are presented as a function of the clustering threshold. Thus figure 5(a) shows clustering performance in terms of  $P_{DE}$  (solid line) and  $P_{CC}$  (dashed line) as a function of  $\eta$  for noise-free signals, and figure 5(b) the corresponding  $I_{\min}$ ,  $\bar{T}$  and  $I_{\max}$ . For  $\eta < 3$ , the number of initiated clusters increases rapidly with decreasing  $\eta$ . Within the interval  $3 < \eta < 7$ , all rhythms initiate at least one cluster, while for  $\eta > 7$ , this is not true for all cases. Removing the "worst case", the above is true for the other cases for  $\eta < 10$ .

The clustering performance is exemplified for one set of parameters in Fig. 6 using  $\eta = 7$ . Correct clustering with minimal number of clusters is achieved for two cases while an extra cluster is initialized for two cases. The reason for the extra cluster is that, for case 2, the temporal search interval is chosen too small for the third morphology from the left resulting in an extra cluster. For case 5 an actual difference in morphologies both on the up and down slopes of the dominant peak is discernible between the first and fourth clusters from the left.

Clustering for a concatenated electrogram is presented for case 3 in Fig. 7, where three distinct rhythm classes can be discerned, viz. normal sinus rhythm followed by supraventricular tachycardia and atrial flutter. The EGM is shown together with the clusters, represented by o, x and +, respectively, assigned for each event. The different rhythms result in three clusters.

Figure 8 is a flow chart illustrating the function in broad outline of an embodiment of the apparatus according to the invention. At block 2 cardiac event features are extracted in the form of wavelet coefficients, and the event is detected, at 4, 6. At block 8 is checked whether the detected event is member of a labelled cluster. If so, the event is added to a class, at 10, and actions associated with that class are performed, at 12.

If the event is not member of a labelled class it is checked if it is a member of an existing cluster, at 14. If so, the event is added to a class, at 16, and it is checked if it is possible to label the cluster, 18, and if so the cluster is labelled, at 20.

If the event is not a member of an existing cluster, block 14, a new cluster is created as described above fitted to the detected cardiac event, at 22.

Thus according to the invention clustering events in the EGM is performed for use in implantable CRM devices, like heart stimulators. The invention is based on feature extraction in the wavelet domain whereupon the

features are clustered based on the Mahalanobis distance criterion. According to advantageous embodiments of the apparatus according to the invention simplifications of the technique is proposed in order to reduce computational complexity to obtain a more implementationally feasible solution.

5 If the apparatus according to the invention is to be used for longer periods of data analysis, large clusters, although old, may be desirable to be kept in some way, while the oldest cluster may be selected to be removed if the application is based on shorter time frames. Also, due to the short data lengths, testing of such algorithms would be of limited value.

10 By combining detector/clusterer with labelling rules of a classifier a complete detector/classifier is obtained with the possibility to more accurate therapy.

15 The labelling need not be done in real time and probably more than one event will be needed to label a cluster. Once the cluster has been labelled using clinical terms, the actions associated with the particular class will be carried out immediately, i.e. in real time. Thus, the rules needed to label the cluster are not used in identifying the event itself.

20 The rules used to label the clusters are based on characteristics of the different possible events. Instead of the exact rules, the characteristics are consequently described.

Figure 9 is a block diagram of an embodiment of an implantable heart stimulator provided with the apparatus according to the invention. Electrodes 30, 32 implanted in the heart 34 of a patient are connected by a lead 36 to an A/D converter 40, via a switch 38 serving as overvoltage protection for the A/D converter 40. In the A/D converter 40 the signal is A/D converted and the digital signal is supplied to a wavelet detector 42.

The detector 42 decides whether a cardiac event is present or not as described earlier. Wavelet coefficients are calculated as well. Parameters of the detector 42 are programmable from the stimulator microprocessor 44. At the 30 detection the coefficients and the RR information are forwarded to the clusterer 46 in which it is determined to which cluster the detected cardiac event belongs, as described previously. The clusterer 46 is preferably of a leader-follower type and also the cluster parameters are programmable from the microprocessor 44.

By the microprocessor 44 suitable therapy is decided depending on assigned cluster for the detected event and possible a priori knowledge about arrhythmia associated with the cluster in question. Thus it is possible to distinguish e.g. ventricular tachycardia from a sinus tachycardia by comparing the parameters  
5 with a known normal sinus rhythm. Parameters of the sinus tachycardia are then supposed to be similar to those of a normal sinus rhythm, whereas parameters of ventricular tachycardia differ significantly.

As an alternative the decision rules can be trained from a number of rhythms and the resulting rules are then used on test data, see Weichao Xu et al.,  
10 "New Bayesian Discriminator for Detection of Atrial Tachyarrhythmias",  
DOI:10.1161/01.CIR.0000012349.14270.54, pp.1472-1479, January 2002. It is then possible to decide that a certain position of the cluster indicates e.g. a sinus rhythm, etc. Also this technique can be based on analysis of the feature vector for a cluster, and it is possible to decide if the beat is broad or narrow, large or small,  
15 or if the rhythm is regular or irregular.

The stimulator in figure 9 also includes a pulse unit 48 with associated battery 50 for delivery of stimulation pulses to the patient's heart 34 depending on the clustering evaluation.

The implantable stimulator shown in figure 9 includes telemetry means 52  
20 with antenna 54 for communication with external equipment, like a programmer.

---